

Fast Moving Horizon State Estimation for Discrete-time Systems Using Single and Multi Iteration Descent Methods

Angelo Alessandri, Mauro Gaggero

Abstract—Descent algorithms based on the gradient, conjugate gradient, and Newton methods are investigated to perform optimization in moving horizon state estimation for discrete-time linear and nonlinear systems. Conditions that ensure the stability of the estimation error are established for single and multi iteration schemes with a least-squares cost function that takes into account only a batch of most recent information. Simulation results show the effectiveness of the proposed approaches also in comparison with techniques based on the Kalman filter.

Index Terms—Moving horizon, State estimation, Gradient method, Conjugate gradient method, Newton method.

I. INTRODUCTION

Moving horizon estimation (MHE) consists in using only a batch of most recent information to estimate the state variables of a dynamic system. At each time instant, the oldest measure is removed from the information batch and replaced with the new one; then, the estimate is obtained by minimizing a least-squares cost function that accounts for the information batch and a prediction about the past state. This paper deals with the reduction of the computational effort by performing only an imperfect optimization at each time instant with the gradient, conjugate gradient, and Newton methods, while guaranteeing the stability of the estimation error.

The first results on MHE date back to [1], where “limited memory” estimation was proposed as an alternative to the Kalman filter. Nowadays, a vast literature on MHE exists, with a number of results for linear and nonlinear systems [2]–[6], large-scale systems [7]–[9], switching systems [10], [11], descriptor systems [12], and systems affected by uncertainties [13]–[16]. The extent of the literature certifies the effectiveness of the MHE methodology and still motivates the investigation of further developments.

Since in general it is necessary to solve the MHE optimization problem at each time instant, the question of completing such a computation within the next time instant is crucial and, when using the estimate in an output feedback loop, even mandatory to apply the control action. However, the reduction of the computational complexity should be pursued without sacrificing stability. Thus, in this paper we address fast moving horizon state estimation for noise-free, discrete-time

systems by performing only a few number of descent steps so as to complete the optimization in real time. Preliminary results concerning the proposed approaches based only on the gradient and Newton methods are reported in [17]. In [18], local stability is proved under quite general, nonlinear assumptions on the system equations for a Newton-based moving horizon state estimator. The use of the Newton method for MHE is explicitly considered in [19].

In the literature on MHE, various approaches are reported on the use of fast techniques to perform the optimization on line [20]–[23]. As compared to such methods, in this paper we investigate the application of single and multi iteration descent algorithms such as the gradient, conjugate gradient, and Newton methods for MHE is presented, and rigorous conditions to ensure the stability of the estimation error are established for both linear and nonlinear systems. Simulation results are presented to compare the proposed approaches, the estimator based on the minimization of the cost function by using a general-purpose optimization tool, and either the Kalman filter (KF) or the extended Kalman filter (EKF) in the linear or nonlinear case, respectively.

The paper is organized as follows. In Section II, the proposed fast approaches to MHE are introduced together with the basic assumptions on the system equations. Sections III, IV, and V report the findings on the stability of the estimation error for the three considered descent algorithms. In Section VI, the stability conditions are compared and discussed. Simulation results are reported in Section VII. Section VIII concludes the paper and prospects future work.

We will adopt the following notation. The symbol (x, y) , where x and y are column vectors, stands for $[x^\top, y^\top]^\top$. The minimum and maximum eigenvalues of a real, symmetric matrix P are denoted by $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$, respectively. Moreover, $P > 0$ ($P \geq 0$) means that P is positive definite (semidefinite). Given a generic matrix M , $|M| := (\lambda_{\max}(M^\top M))^{1/2} = (\lambda_{\max}(MM^\top))^{1/2}$. Accordingly, in the special case of a vector v , $|v| := (v^\top v)^{1/2}$ denotes its Euclidean norm. Any smooth real function that is differentiated at least n times and having a continuous derivative of order n is said of class C^n . The gradient of a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a $n \times m$ matrix where the i -th column is given by the gradient $\nabla f_i(x)$ of the i -th function, i.e.,

$$\nabla f(x) := [\nabla f_1(x), \nabla f_2(x), \dots, \nabla f_m(x)] .$$

A. Alessandri is with the Department of Mechanical Engineering, University of Genoa, P.le Kennedy Pad. D, 16129 Genoa, Italy e-mail: alessandri@time.unige.it.

M. Gaggero is with the Institute of Intelligent Systems for Automation, National Research Council of Italy, Via De Marini 6, 16149 Genoa, Italy e-mail: mauro.gaggero@cnr.it.

Given a smooth function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, its Hessian matrix is denoted by $\nabla^2 g(x)$.

II. MOVING HORIZON STATE ESTIMATION

Consider the discrete-time dynamic system

$$x_{t+1} = f(x_t, u_t) \quad (1a)$$

$$y_t = h(x_t, u_t) \quad (1b)$$

where $t = 0, 1, \dots$ is the time instant, $x_t \in \mathbb{R}^n$ is the state vector, $u_t \in \mathbb{R}^p$ is the control vector, and $y_t \in \mathbb{R}^m$ is the output vector.

An MHE strategy for (1) relies on the information obtained in the recent past by computing a state estimate of x_t at the current time t , denoted by $\hat{x}_{t|t}$, based only on the information given by $y_{t-N}, \dots, y_t, u_{t-N}, \dots, u_t$, where N is the length of the window. In addition, we consider also a “prediction” \bar{x}_{t-N} of the state x_{t-N} at the beginning of the moving window. Thus, let us consider a least-squares cost function

$$J_t(\hat{x}_{t-N|t}) = \mu |\hat{x}_{t-N|t} - \bar{x}_{t-N}|^2 + \sum_{i=t-N}^t |y_i - h(\hat{x}_{i|t}, u_i)|^2 \quad (2)$$

where $\mu \geq 0$ and we denote the estimates of x_{t-N}, \dots, x_{t-1} at time t by $\hat{x}_{t-N|t}, \dots, \hat{x}_{t-1|t}$, respectively.

At each time t , the cost (2) has to be minimized together with the constraints

$$\hat{x}_{i+1|t} = f(\hat{x}_{i|t}, u_i), \quad i = t-N, \dots, t-1. \quad (3)$$

The estimate of x_{t-N} at time t is the minimizer of (2), i.e.,

$$\hat{x}_{t-N|t} = \operatorname{argmin} \{ J_t(\hat{x}), \hat{x} \in \mathbb{R}^n \text{ subject to (3)} \}. \quad (4)$$

The remaining estimates at time t descend from (3) by propagating $\hat{x}_{t-N|t}$ forward via the state equation as follows: $\hat{x}_{t-N+i+1|t} = f(\hat{x}_{t-N+i|t}, u_{t-N+i})$, $i = 0, \dots, N-1$. Moreover, we need to assign \bar{x}_{t-N} , for which various choices can be made. In particular, we take the prediction that results from the previous estimation step, i.e., we let $\bar{x}_{t-N} = f(\hat{x}_{t-N-1|t-1}, u_{t-N-1})$. As to the initialization, some a-priori prediction of x_0 , denoted by \bar{x}_0 , has to be chosen at the beginning. Summing up, the three main steps of the proposed MHE procedure are *optimization*, *propagation*, and *prediction*, as pictorially shown in Fig. 1.

In line with preliminary results presented in [17], we deal with MHE based on the gradient, conjugate gradient, and Newton methods. More specifically, we renounce to perform the “full” minimization as in (4) and, instead, we compute the estimate of the state by applying only one or more descent steps. Clearly, such an imperfect optimization entails a reduced computational burden but, under suitable assumptions, guarantees the stability of the estimation error, as will be proved in the following.

First, we will investigate the use of the gradient method with one descent step at time $t = N, N+1, \dots$ as follows:

$$\hat{x}_{t-N|t} = \bar{x}_{t-N|t} - \alpha \nabla J_t(\bar{x}_{t-N|t}), \quad t = N, N+1, \dots \quad (5)$$

where $\alpha > 0$ is the step size.

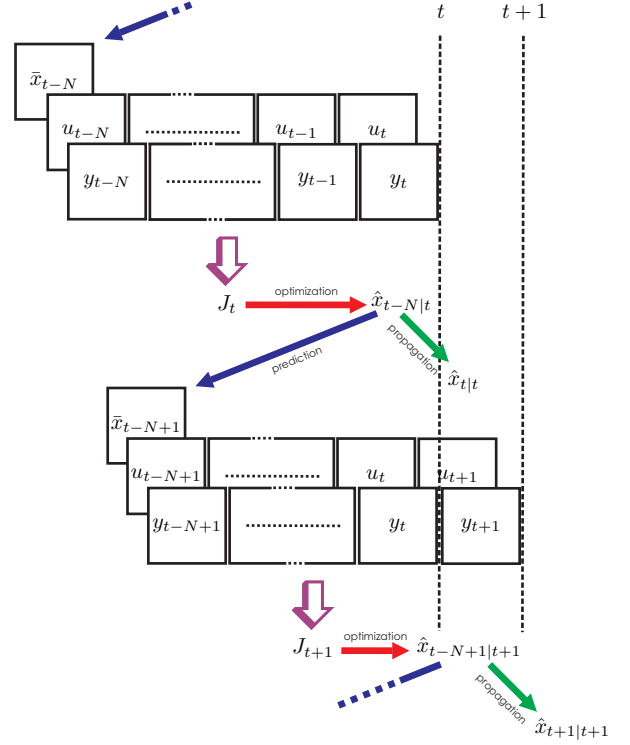


Fig. 1. Pictorial description of the MHE approach.

The second approach we will consider is inspired¹ by the conjugate gradient method. It is composed of the following two steps at time $t = N, N+1, \dots$:

$$\tilde{x}_{t-N|t} = \bar{x}_{t-N|t} - \alpha \nabla J_t(\bar{x}_{t-N|t}) \quad (6a)$$

$$\hat{x}_{t-N|t} = \tilde{x}_{t-N|t} + \alpha \left(-\nabla J_t(\tilde{x}_{t-N|t}) + \frac{\nabla J_t(\tilde{x}_{t-N|t})^\top \nabla J_t(\tilde{x}_{t-N|t})}{\nabla J_t(\bar{x}_{t-N|t})^\top \nabla J_t(\bar{x}_{t-N|t})} \nabla J_t(\bar{x}_{t-N|t}) \right) \quad (6b)$$

where $\alpha > 0$ is the step size. In the first step we get the preliminary estimate $\tilde{x}_{t-N|t}$ using a gradient descent iteration. In the case $\nabla J_t(\bar{x}_{t-N|t}) = 0$, the final estimate $\hat{x}_{t-N|t}$ is chosen equal to $\tilde{x}_{t-N|t}$. If $\nabla J_t(\bar{x}_{t-N|t}) \neq 0$, we compute another descent iteration along a conjugate direction using the Fletcher-Reeves formula [25].

Finally, the application of one iteration step of the Newton method at time $t = N, N+1, \dots$ provides the following:

$$\hat{x}_{t-N|t} = \bar{x}_{t-N|t} - (\nabla^2 J_t(\bar{x}_{t-N|t}))^{-1} \nabla J_t(\bar{x}_{t-N|t}). \quad (7)$$

Of course, the estimation procedures (5), (6), and (7) need to be completed with propagation and prediction, as previously pointed out. We will refer to (5), (6), and (7) as gradient, conjugate gradient, and Newton moving horizon state estimators or GMHE, CGMHE, and NMHE for short, respectively.

The stability of the estimation errors for the proposed schemes requires some common assumptions we will briefly

¹The classical conjugate gradient method (see [24, p. 130]) requires to perform a line search by selecting the optimal step size at each iteration, whereas the step size of our approach is fixed.

introduce in the following. Later on, in the next sections we will show the stability properties of the GMHE, CGMHE, and NMHE separately. Toward this end, let $x_t \in X \subset \mathbb{R}^n$ and $u_t \in U \subset \mathbb{R}^p$, where X and U are compact sets. Such assumptions on X and U may be relaxed in the treatment of the problem in the linear case, where $X \equiv \mathbb{R}^n$ and $U \equiv \mathbb{R}^p$.

The application of the descent steps (5), (6), and (7) does not ensure that the estimate $\hat{x}_{t|t}$ resulting from the propagation of $\hat{x}_{t-N|t}$ belong to X . To enforce such a constraint, one may apply a projection on X , i.e.,

$$\hat{x}_{t|t} = \text{Proj}_X (f^{u_{t-1}} \circ f^{u_{t-2}} \circ \dots \circ f^{u_{t-N}} (\hat{x}_{t-N|t})) \quad (8)$$

where, from now on, to avoid burdening the notation we will write $f^u(\cdot)$ and $h^u(\cdot)$ instead of $f(\cdot, u)$ and $h(\cdot, u)$, respectively. In general, the projection is obtained by the minimization of a distance cost function, which is just what here is avoided by performing only one descent step instead of a full optimization. However, simple forms of the set X such as rectangles or polytopes require a much less computational effort, which can be accomplished in quite a quick way. Of course, the application of the projection provides “better” estimates in terms of a lower norm of the estimation error, thus the projected estimates will be simply dropped in the proofs given later.

Owing to the projection (8), all the state estimates belong to X . Moreover, we have to assume that any convex combination of such estimates is inside X or, in other words, to deal with the convex hull of X . Thus, to simplify the notation, from now on we will refer still to X but intend its convex hull.

The single iteration methods (5), (6), and (7) can be easily extended to a multi iteration setup. For the sake of brevity, let us consider only the extension of the NMHE with $M > 1$ iterations as shown in Fig. 2. In particular, the descent steps of the inner loop allow for an increase of the speed of convergence, as will be clarified in Section VI. We will refer to such an approach as NMHE- M , and call similarly GMHE- M and CGMHE- M the GMHE and CGMHE with M iterations, respectively. Finally, it is worth noting the relation of the proposed approaches with the so-called inexact Newton methods [26], [27].

Multi iteration MHE based on the Newton method	
Input: $\bar{x}_{t-N t}$, $\nabla J_t(\cdot)$, $\nabla^2 J_t(\cdot)$, M	
Output: $\hat{x}_{t t}$	
1: $\xi_{t-N}(1) \leftarrow \bar{x}_{t-N t}$	
2: for m from 1 to M do	
$\hat{\xi}_{t-N}(m+1) \leftarrow \hat{\xi}_{t-N}(m) - \left(\nabla^2 J_t(\hat{\xi}_{t-N}(m)) \right)^{-1}$	
$\times \nabla J_t(\hat{\xi}_{t-N}(m))$	
3: end	
4: $\hat{x}_{t-N t} \leftarrow \hat{\xi}_{t-N}(M+1)$	
5: $\hat{x}_{t t} \leftarrow \text{Proj}_X (f^{u_{t-1}} \circ f^{u_{t-2}} \circ \dots \circ f^{u_{t-N}} (\hat{x}_{t-N t}))$	

Fig. 2. Sketch of a multi iteration MHE scheme.

For the sake of brevity, the analysis of stability of the estimation error will be detailed only in the single iteration case for GMHE, CGMHE, and NMHE. Toward this end, we need to assume the following.

Assumption 1: The functions f and h are of class C^2 with Lipschitz continuous Hessian matrices. Specifically, f satisfies the Lipschitz condition

$$|f^u(x') - f^u(x'')| \leq k_f |x' - x''|, \quad \forall x', x'' \in X$$

for some $k_f > 0$ and independently of $u \in U$. ■

From now on, let $y_{t-N}^t := (y_t, y_{t-1}, \dots, y_{t-N})$, $u_{t-N}^t := (u_t, u_{t-1}, \dots, u_{t-N})$, and $H : X \times U^{N+1} \rightarrow \mathbb{R}^q$ defined as follows:

$$H(x_{t-N}, u_{t-N}^t) := \begin{pmatrix} h^{u_t} \circ f^{u_{t-1}} \circ f^{u_{t-2}} \circ \dots \circ f^{u_{t-N}}(x_{t-N}) \\ h^{u_{t-1}} \circ f^{u_{t-2}} \circ \dots \circ f^{u_{t-N}}(x_{t-N}) \\ \vdots \\ h^{u_{t-N+1}} \circ f^{u_{t-N}}(x_{t-N}) \\ h^{u_{t-N}}(x_{t-N}) \end{pmatrix}$$

where $q := (N+1)m$ and S^k denotes the Cartesian product of k sets S . Thus, the cost (2) can be written as follows:

$$J_t(\hat{x}) = \mu |\hat{x} - \bar{x}_{t-N|t}|^2 + |y_{t-N}^t - H(\hat{x}, u_{t-N}^t)|^2. \quad (9)$$

The gradient and the Hessian of the cost function (9) w.r.t. the argument \hat{x} can be computed as follows:

$$\begin{aligned} \nabla J_t(\hat{x}) &= 2\mu (\hat{x} - \bar{x}_{t-N|t}) - 2 \sum_{i=1}^q \bar{y}_i \nabla H_i(\hat{x}, u_{t-N}^t) \\ &\quad + 2 \sum_{i=1}^q H_i(\hat{x}, u_{t-N}^t) \nabla H_i(\hat{x}, u_{t-N}^t) \end{aligned}$$

and

$$\begin{aligned} \nabla^2 J_t(\hat{x}) &= 2\mu I - 2 \sum_{i=1}^q \bar{y}_i \nabla^2 H_i(\hat{x}, u_{t-N}^t) \\ &\quad + 2 \sum_{i=1}^q H_i(\hat{x}, u_{t-N}^t) \nabla^2 H_i(\hat{x}, u_{t-N}^t) \\ &\quad + 2 \sum_{i=1}^q \nabla H_i(\hat{x}, u_{t-N}^t) \nabla H_i(\hat{x}, u_{t-N}^t)^\top \end{aligned} \quad (10)$$

where $y_{t-N}^t =: \bar{y} \in Y^{N+1}$, \bar{y}_i is the i -th component of \bar{y} , and $Y := h(X, U)$.

Thanks to Assumption 1, there exists $k_J > 0$ such that the Lipschitz inequality

$$|\nabla^2 J_t(x') - \nabla^2 J_t(x'')| \leq k_J |x' - x''|, \quad \forall x', x'' \in X \quad (11)$$

holds for all t and with $k_J > 0$ that does not depend on μ .

Now we consider the Hessian matrix (10) in more detail by rewriting it as follows:

$$\nabla^2 J_t(\hat{x}) = 2\mu I + 2 R(\hat{x}, \bar{y}, \bar{u}) \quad (12)$$

where $u_{t-N}^t =: \bar{u} \in U^{N+1}$ and $R : X \times Y^{N+1} \times U^{N+1} \rightarrow S_n(\mathbb{R}) := \{P \in \mathbb{R}^{n \times n} : P = P^\top\}$ is defined as follows

$$\begin{aligned} R(\hat{x}, \bar{y}, \bar{u}) &:= - \sum_{i=1}^q \bar{y}_i \nabla^2 H_i(\hat{x}, \bar{u}) + \sum_{i=1}^q H_i(\hat{x}, \bar{u}) \nabla^2 H_i(\hat{x}, \bar{u}) \\ &\quad + \sum_{i=1}^q \nabla H_i(\hat{x}, \bar{u}) \nabla H_i(\hat{x}, \bar{u})^\top. \end{aligned}$$

The definition above allows to formulate an assumption of observability on (1) as follows (see [28] for details on the various observability definitions and their connections).

Assumption 2: Let the system (1) be such that

$$\delta_{\min} := \min_{\hat{x} \in X, \bar{y} \in Y^{N+1}, \bar{u} \in U^{N+1}} \lambda_{\min}(R(\hat{x}, \bar{y}, \bar{u})) \quad (13)$$

is strictly positive. ■

Moreover, let

$$\delta_{\max} := \max_{\hat{x} \in X, \bar{y} \in Y^{N+1}, \bar{u} \in U^{N+1}} \lambda_{\max}(R(\hat{x}, \bar{y}, \bar{u})) . \quad (14)$$

Concerning the proofs of stability that will be discussed in the next sections, it is worth noting that it is sufficient to prove that the estimation error $e_{t-N} := x_{t-N} - \hat{x}_{t-N|t}$ is exponentially stable since the inequality

$$|x_t - \hat{x}_{t|t}| \leq (k_f)^N |e_{t-N}|$$

holds owing to Assumption 1. To avoid burdening the notation, the stability proofs in the following will be presented by referring to an autonomous system setting, i.e., we will use $f(\cdot)$ instead of $f^u(\cdot)$. The extension to the general case is straightforward.

III. MHE BASED ON GRADIENT METHOD

In this section, first we will address the stability properties of the estimation error given by the GMHE with a suitable choice of the descent step size α in the general case for a nonlinear system (1), and then with linear dynamic and measurement equations.

Under the general system setting (1), we state the following.

Theorem 1: The estimation error for system (1) given by the GMHE (5) with a constant step size $\alpha > 0$ such that

$$\max(|1 - 2\alpha\delta_{\min}|, |1 - 2\alpha\delta_{\max}|) k_f < 1 \quad (15)$$

is exponentially stable.

Proof. The estimate based on the GMHE at time $t+1$ is

$$\hat{x}_{t-N+1|t+1} = \bar{x}_{t-N+1|t+1} - \alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})$$

and hence we easily get

$$\begin{aligned} e_{t-N+1} &= x_{t-N+1} - \hat{x}_{t-N+1|t+1} = f(x_{t-N}) \\ &\quad - f(\hat{x}_{t-N|t}) + \alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}). \end{aligned} \quad (16)$$

Using a Taylor expansion centered in x_{t-N+1} with exact remainder term, it follows that there exists $\lambda \in [0, 1]$ such that

$$\begin{aligned} \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) &= \nabla J_{t+1}(x_{t-N+1}) \\ &\quad + \nabla^2 J_{t+1}(\lambda x_{t-N+1} + (1-\lambda)\bar{x}_{t-N+1|t+1}) (\bar{x}_{t-N+1|t+1} \\ &\quad - x_{t-N+1}), \end{aligned}$$

and therefore, after some simplifications,

$$\begin{aligned} \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) &= 2\mu (f(x_{t-N}) - f(\hat{x}_{t-N|t})) \\ &\quad + \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) (f(\hat{x}_{t-N|t}) \\ &\quad - f(x_{t-N})). \end{aligned} \quad (17)$$

Using (17), (16) yields

$$\begin{aligned} e_{t-N+1} &= \left((1 + 2\alpha\mu)I - \alpha \nabla^2 J_{t+1}(\lambda f(x_{t-N}) \right. \\ &\quad \left. + (1-\lambda)f(\hat{x}_{t-N|t})) \right) (f(x_{t-N}) - f(\hat{x}_{t-N|t})). \end{aligned} \quad (18)$$

Based on (13) and (14), from (12) we obtain the matrix inequalities

$$\begin{aligned} 2(\mu + \delta_{\min})I &\leq \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) \\ \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) &\leq 2(\mu + \delta_{\max})I \end{aligned}$$

and hence

$$\begin{aligned} &\left| (1 + 2\alpha\mu)I - \alpha \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) \right| \\ &\leq \max(|1 - 2\alpha\delta_{\min}|, |1 - 2\alpha\delta_{\max}|). \end{aligned} \quad (19)$$

Using (19) and Assumption 1, from (18) it follows that

$$\begin{aligned} |e_{t-N+1}| &\leq \left| (1 + 2\alpha\mu)I - \alpha \nabla^2 J_{t+1}(\lambda f(x_{t-N}) \right. \\ &\quad \left. + (1-\lambda)f(\hat{x}_{t-N|t})) \right| |f(x_{t-N}) - f(\hat{x}_{t-N|t})| \\ &\leq \max(|1 - 2\alpha\delta_{\min}|, |1 - 2\alpha\delta_{\max}|) k_f |x_{t-N} - \hat{x}_{t-N|t}|. \end{aligned}$$

If we apply the bound above at the first step and for $t = N, N+1, \dots$, we obtain

$$\begin{aligned} |e_{t-N+1}| &\leq (\max(|1 - 2\alpha\delta_{\min}|, |1 - 2\alpha\delta_{\max}|) k_f)^{t-N+2} \\ &\quad \times |x_0 - \bar{x}_0| \end{aligned}$$

and therefore the exponential stability of the estimation error is ensured if (15) holds. ■

Note that the stability requires to choose α independent of μ since the gradient of the cost that corresponds to the prediction has the first term null and thus it is not affected by μ .

Now, let us focus on the linear system setting given by

$$x_{t+1} = A x_t + B u_t \quad (20a)$$

$$y_t = C x_t + D u_t \quad (20b)$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{m \times n}$, and $D \in \mathbb{R}^{m \times p}$.

It is straightforward to show that (20) can replace (1) in the proof of Theorem 1 without any conceptual difficulty. Thus, we have the following.

Corollary 1: The estimation error for system (20) given by the GMHE (5) with a constant step size $\alpha > 0$ such that

$$\max(|1 - 2\alpha\delta_{\min}|, |1 - 2\alpha\delta_{\max}|) |A| < 1$$

is exponentially stable.

Proof. It is in line with the proof of Theorem 1 with $|A|$ instead of k_f . ■

IV. MHE BASED ON CONJUGATE GRADIENT METHOD

The stability of the estimation error provided by the CGMHE in performing estimation for nonlinear dynamic systems like (1) can be ensured under suitable conditions as follows.

Theorem 2: The estimation error for system (1) given by the CGMHE (6) with a constant step size $\alpha > 0$ and $\mu \geq 0$ is exponentially stable if

$$2k_f \delta_{\max} \max \left\{ \left| -1/(2\delta_{\max}) + 2\alpha - 2\alpha^2(\mu + \delta_{\max}) - \alpha\gamma(\alpha, \mu) \right|, \left| -1/(2\delta_{\min}) + 2\alpha - 2\alpha^2(\mu + \delta_{\min}) \right| \right\} < 1 \quad (21)$$

where

$$\gamma(\alpha, \mu) := \begin{cases} (1 - 2\alpha(\mu + \delta_{\min}))^2, & \text{if } \alpha \in [0, 1/(2\mu + \delta_{\min} + \delta_{\max})], \\ (2\alpha(\mu + \delta_{\max}) - 1)^2, & \text{if } \alpha > 1/(2\mu + \delta_{\min} + \delta_{\max}). \end{cases} \quad (22)$$

Proof. From (6), it follows that the estimation error given by the CGMHE at time $t + 1$ is

$$e_{t-N+1} = x_{t-N+1} - \hat{x}_{t-N+1|t+1} = f(x_{t-N}) - f(\hat{x}_{t-N|t}) + \alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) - \alpha (-\nabla J_{t+1}(\tilde{x}_{t-N+1|t+1}) + K(\tilde{x}_{t-N+1|t+1}, \bar{x}_{t-N+1|t+1}) \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})) \quad (23)$$

where

$$K(\tilde{x}_{t-N+1|t+1}, \bar{x}_{t-N+1|t+1}) := \frac{\nabla J_{t+1}(\tilde{x}_{t-N+1|t+1})^\top \nabla J_{t+1}(\tilde{x}_{t-N+1|t+1})}{\nabla J_{t+1}(\bar{x}_{t-N+1|t+1})^\top \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})} I. \quad (24)$$

Let focus on (23). Likewise in the proof of Theorem 1, we use a Taylor expansion of $\nabla J_{t+1}(\bar{x}_{t-N+1|t+1})$ centered in x_{t-N+1} with exact remainder term, i.e.,

$$\nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) = 2\mu (f(x_{t-N}) - f(\hat{x}_{t-N|t})) + H_1^J(x_{t-N}, \hat{x}_{t-N|t}) (f(\hat{x}_{t-N|t}) - f(x_{t-N})) \quad (25)$$

where

$$H_1^J(x_{t-N}, \hat{x}_{t-N|t}) := \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t}))$$

for some $\lambda \in [0, 1]$. Moreover, we use a Taylor expansion of $\nabla J_{t+1}(\tilde{x}_{t-N+1|t+1})$ centered in \bar{x}_{t-N+1} . Specifically, from the first equation of (6) we get

$$\begin{aligned} \nabla J_{t+1}(\tilde{x}_{t-N+1|t+1}) &= \nabla J_{t+1}(\bar{x}_{t-N+1|t+1} - \alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})) \\ &= \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) + H_2^J(\bar{x}_{t-N+1|t+1}, \alpha) (-\alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})) \\ &= (I - \alpha H_2^J(\bar{x}_{t-N+1|t+1}, \alpha)) \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) \end{aligned} \quad (26)$$

where

$$H_2^J(\bar{x}_{t-N+1|t+1}, \alpha) := \nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1} - (1-\tilde{\lambda})\alpha \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}))$$

for some $\tilde{\lambda} \in [0, 1]$. Using (25) and (26), from (23) it follows that

$$e_{t-N+1} = \left((2\mu I - H_1^J)^{-1} + 2\alpha I - \alpha^2 H_2^J - \alpha K \right) (2\mu I - H_1^J) (f(x_{t-N}) - f(\hat{x}_{t-N|t})) \quad (27)$$

where the functional dependences of H_1^J , H_2^J , and K on their arguments are dropped from now on to avoid burdening the

notation. Again, we need to derive some matrix bounds to deal with (27).

Since (12) yields

$$2(\mu + \delta_{\min})I \leq H_i^J \leq 2(\mu + \delta_{\max})I \quad (28)$$

for $i = 1, 2$, we have

$$-1/(2\delta_{\max})I \leq (2\mu I - H_1^J)^{-1} \leq -1/(2\delta_{\min})I \quad (29)$$

$$-2\alpha^2(\mu + \delta_{\max})I \leq -\alpha^2 H_2^J \leq -2\alpha^2(\mu + \delta_{\min})I \quad (30)$$

$$|2\mu I - H_1^J| \leq 2\delta_{\max}. \quad (31)$$

Using (26), from (24) we obtain

$$K = \frac{\nabla J^\top (I - \alpha H_2^J)^\top (I - \alpha H_2^J) \nabla J}{|\nabla J|^2} I$$

where $\nabla J := \nabla J_{t+1}(\bar{x}_{t-N+1|t+1})$ for the sake of brevity, and thus

$$|K| \leq |I - \alpha H_2^J|^2. \quad (32)$$

In order to bound the r.h.s. of (32), from (28) we have

$$(1 - 2\alpha(\mu + \delta_{\max}))I \leq I - \alpha H_2^J \leq (1 - 2\alpha(\mu + \delta_{\min}))I$$

and then

$$|I - \alpha H_2^J| \leq \max(|1 - 2\alpha(\mu + \delta_{\min})|, |1 - 2\alpha(\mu + \delta_{\max})|). \quad (33)$$

If we expand the r.h.s. of (33) and use (32), we obtain

$$|K| \leq |I - \alpha H_2^J|^2 \leq \gamma(\alpha, \mu)$$

where $\gamma(\alpha, \mu)$ is defined in (22). Thus, we can write

$$-\alpha\gamma(\alpha, \mu)I \leq -\alpha K \leq 0. \quad (34)$$

From (29), (30), and (34), it follows that

$$\begin{aligned} &(-1/(2\delta_{\max}) + 2\alpha - 2\alpha^2(\mu + \delta_{\max}) - \alpha\gamma(\alpha, \mu))I \\ &\leq (2\mu I - H_1^J)^{-1} + 2\alpha I - \alpha^2 H_2^J - \alpha K \\ &\leq (-1/(2\delta_{\min}) + 2\alpha - 2\alpha^2(\mu + \delta_{\min}))I \end{aligned}$$

and, as a consequence,

$$\begin{aligned} &\left| (2\mu I - H_1^J)^{-1} + 2\alpha I - \alpha^2 H_2^J - \alpha K \right| \\ &\leq \max \left\{ \left| -1/(2\delta_{\max}) + 2\alpha - 2\alpha^2(\mu + \delta_{\max}) - \alpha\gamma(\alpha, \mu) \right|, \left| -1/(2\delta_{\min}) + 2\alpha - 2\alpha^2(\mu + \delta_{\min}) \right| \right\}. \end{aligned} \quad (35)$$

Using (31), (35), and Assumption 1, (27) yields

$$|e_{t-N+1}| \leq 2k_f \delta_{\max} \max \left\{ \left| -1/(2\delta_{\max}) + 2\alpha - 2\alpha^2(\mu + \delta_{\max}) - \alpha\gamma(\alpha, \mu) \right|, \left| -1/(2\delta_{\min}) + 2\alpha - 2\alpha^2(\mu + \delta_{\min}) \right| \right\} |e_{t-N}|$$

and thus we get exponential stability if (21) holds. \square

Let us now move to the linear system setting (20). It is easy to prove the following.

Corollary 2: The estimation error for system (20) given by the GMHE (5) with a constant step size $\alpha > 0$ and $\mu \geq 0$ is exponentially stable if

$$2|A| \delta_{\max} \max \left\{ \left| -1/(2\delta_{\max}) + 2\alpha - 2\alpha^2(\mu + \delta_{\max}) - \alpha\gamma(\alpha, \mu) \right|, \left| -1/(2\delta_{\min}) + 2\alpha - 2\alpha^2(\mu + \delta_{\min}) \right| \right\} < 1. \quad \square$$

V. MHE BASED ON NEWTON METHOD

As regards the NMHE with the general nonlinear system setup given by (1), the following local stability result holds.

Theorem 3: The estimation error for system (1) given by the NMHE (7) is exponentially stable if

$$\alpha_0 < 1 \quad (36)$$

where

$$\alpha_0 := \frac{(k_f k_J |x_0 - \bar{x}_0| + 2\mu) k_f}{2(\mu + \delta_{\min})}. \quad (37)$$

Proof. The NMHE estimate update at time $t + 1$ is

$$\begin{aligned} \hat{x}_{t-N+1|t+1} &= \bar{x}_{t-N+1|t+1} - (\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}))^{-1} \\ &\quad \times \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}). \end{aligned}$$

Hence, we obtain

$$\begin{aligned} e_{t-N+1} &= x_{t-N+1} - \hat{x}_{t-N+1|t+1} = x_{t-N+1} - \bar{x}_{t-N+1|t+1} \\ &\quad + (\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}))^{-1} \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) \\ &= f(x_{t-N}) - f(\hat{x}_{t-N|t}) + (\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}))^{-1} \\ &\quad \times \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}). \end{aligned} \quad (38)$$

After using a Taylor expansion centered in x_{t-N+1} with exact remainder term, we have

$$\begin{aligned} \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) &= \nabla J_{t+1}(x_{t-N+1}) + \nabla^2 J_{t+1}(\lambda x_{t-N+1} \\ &\quad + (1-\lambda)\bar{x}_{t-N+1|t+1}) (\bar{x}_{t-N+1|t+1} - x_{t-N+1}) \end{aligned}$$

for some $\lambda \in [0, 1]$, and therefore

$$\begin{aligned} \nabla J_{t+1}(\bar{x}_{t-N+1|t+1}) &= 2\mu (f(x_{t-N}) - f(\hat{x}_{t-N|t})) \\ &\quad + \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) (f(\hat{x}_{t-N|t}) \\ &\quad - f(x_{t-N})). \end{aligned} \quad (39)$$

Using (39), from (38) it follows that

$$\begin{aligned} e_{t-N+1} &= f(x_{t-N}) - f(\hat{x}_{t-N|t}) + (\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}))^{-1} \\ &\quad \times \left(2\mu (f(x_{t-N}) - f(\hat{x}_{t-N|t})) + \nabla^2 J_{t+1}(\lambda f(x_{t-N}) \right. \\ &\quad \left. + (1-\lambda)f(\hat{x}_{t-N|t})) (f(\hat{x}_{t-N|t}) - f(x_{t-N})) \right) \\ &= (\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}))^{-1} \left(\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1}) \right. \\ &\quad \left. - \nabla^2 J_{t+1}(\lambda f(x_{t-N}) + (1-\lambda)f(\hat{x}_{t-N|t})) + 2\mu I \right) \\ &\quad \times (f(x_{t-N}) - f(\hat{x}_{t-N|t})). \end{aligned} \quad (40)$$

From (12) and thanks to Assumption 2, we obtain

$$\nabla^2 J_t(\bar{x}_{t-N+1|t+1}) \geq 2(\mu + \delta_{\min})I$$

and hence

$$|\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1})^{-1}| \leq \frac{1}{2(\mu + \delta_{\min})}. \quad (41)$$

Using (11) and (41), from (40) it follows that

$$\begin{aligned} |e_{t-N+1}| &\leq |\nabla^2 J_{t+1}(\bar{x}_{t-N+1|t+1})^{-1}| |\nabla^2 J_{t+1}(f(\hat{x}_{t-N|t})) \\ &\quad - \nabla^2 J_{t+1}(\lambda f(x_{t-N}) \\ &\quad + (1-\lambda)f(\hat{x}_{t-N|t})) + 2\mu I| |f(x_{t-N}) - f(\hat{x}_{t-N|t})| \\ &\leq \frac{(k_J k_f \lambda |x_{t-N} - \hat{x}_{t-N|t}| + 2\mu) k_f}{2(\mu + \delta_{\min})} |x_{t-N} - \hat{x}_{t-N|t}| \\ &\leq \frac{(k_J k_f |e_{t-N}| + 2\mu) k_f}{2(\mu + \delta_{\min})} |e_{t-N}| = \alpha_{t-N+1} |e_{t-N}| \end{aligned}$$

also since $\lambda \in [0, 1]$, where

$$\alpha_{t-N+1} := \frac{(k_J k_f |e_{t-N}| + 2\mu) k_f}{2(\mu + \delta_{\min})}$$

for $t = N, N+1, \dots$. However, notice that at the first step we have

$$|e_0| \leq \alpha_0 |x_0 - \bar{x}_0|$$

with α_0 defined as in (37) and

$$\alpha_1 := \frac{(k_J k_f |e_0| + 2\mu) k_f}{2(\mu + \delta_{\min})} \leq \alpha_0$$

thanks to (36). Thus, it follows that $\alpha_1 < 1$ and

$$|e_1| \leq \alpha_1 |e_0| \leq \alpha_0 |e_0| \leq \alpha_0^2 |x_0 - \bar{x}_0|.$$

If we proceed by induction, it is straightforward to get

$$|e_{t-N+1}| \leq \alpha_0^{t-N+2} |x_0 - \bar{x}_0|.$$

and then conclude about exponential stability. \square

The stability of the NMHE can be proved to be global under linear assumptions. Toward this end, in the following we analyze the NMHE for the linear system (20).

Corollary 3: The estimation error for system (20) given by the NMHE (7) with $\mu \geq 0$ such that

$$\frac{\mu |A|}{\mu + \delta_{\min}} < 1 \quad (42)$$

is exponentially stable, and

$$|e_{t-N}| \leq \left(\frac{\mu |A|}{\mu + \delta_{\min}} \right)^{t-N+1} |x_0 - \bar{x}_0|. \quad (43)$$

Proof. First of all, notice that

$$x_{\tau+i} = A^i x_\tau + \sum_{j=0}^{i-1} A^j B u_{\tau+i-j-1},$$

with $\tau \geq 0$ and $i \geq 1$ both integer. Second, we need to compute the gradient and Hessian of the cost function, which is given by

$$\begin{aligned} J_{t+1}(\hat{x}_{t-N+1|t+1}) &= \mu |\hat{x}_{t-N+1|t+1} - \bar{x}_{t-N+1}|^2 \\ &\quad + \sum_{i=0}^N |y_{t-N+1+i} - C \hat{x}_{t-N+1+i|t+1} - D u_{t-N+1+i}|^2 \end{aligned} \quad (44)$$

where

$$\hat{x}_{t-N+1+i|t+1} = A^i \hat{x}_{t-N+1|t+1} + \sum_{j=0}^{i-1} A^j B u_{t-N+1+i-j} \quad (45)$$

for any integer $i \geq 1$. Using (45) in (44), we obtain

$$J_{t+1}(\hat{x}_{t-N+1|t+1}) = \mu |\hat{x}_{t-N+1|t+1} - \bar{x}_{t-N+1}|^2 + \sum_{i=0}^N \left| y_{t-N+1+i} - C A^i \hat{x}_{t-N+1|t+1} - C \sum_{j=0}^{i-1} A^j B u_{t-N+1-j} - D u_{t-N+1+i} \right|^2$$

and

$$\begin{aligned} \nabla J_{t+1}(\hat{x}_{t-N+1|t+1}) &= 2\mu (\hat{x}_{t-N+1|t+1} - \bar{x}_{t-N+1}) \\ &+ 2 \sum_{i=0}^N \left((C A^i)^\top (C A^i) \hat{x}_{t-N+1|t+1} - (C A^i)^\top \right. \\ &\quad \times \left. \left(y_{t-N+1+i} - C \sum_{j=0}^{i-1} A^j B u_{t-N+1-j} - D u_{t-N+1+i} \right) \right) \\ \nabla^2 J_{t+1}(\hat{x}_{t-N+1|t+1}) &= 2 \left(\mu I + \sum_{i=0}^N (C A^i)^\top (C A^i) \right). \end{aligned}$$

Based on the aforesaid, after a little algebra, we get

$$e_{t-N+1} = \mu \left(\mu I + \sum_{i=0}^N (C A^i)^\top (C A^i) \right)^{-1} A e_{t-N}. \quad (46)$$

Notice that Assumption 2 turns into a condition of full rank on the observability matrix $(C^\top, (CA)^\top, \dots, (CA^N)^\top)^\top$.

In line with the proof of Theorem 3, if

$$\delta_{\min} := \lambda_{\min} \left(\sum_{i=0}^N (C A^i)^\top (C A^i) \right) > 0,$$

we have

$$\left(\mu I + \sum_{i=0}^N (C A^i)^\top (C A^i) \right)^{-1} < \frac{1}{\mu + \delta_{\min}} I$$

and hence

$$\left| \mu \left(\mu I + \sum_{i=0}^N (C A^i)^\top (C A^i) \right)^{-1} A \right| \leq \frac{\mu |A|}{\mu + \delta_{\min}}. \quad (47)$$

Using (47), from (46) we obtain (43), thus concluding the proof. \square

As compared to Theorem 3, it is worth noting that the stability result in Corollary 3 is global with a suitable choice of μ .

VI. STABILITY CONDITIONS AND CONVERGENCE RATES

In this section, the stability results reported so far for all the proposed MHE single iteration approaches are discussed and extended to multi iteration schemes, which ensure a higher precision but require an increased computational effort.

First of all, note that Theorem 3 ensures only local stability for the Newton-based MHE in the nonlinear case because of a condition that depends on the initial estimation error. By contrast, the stability of the GMHE and CGMHE in Theorems 1 and 2 is of global type, but it is ensured only for a small value of k_f . Such results turn out to be novel as compared

to the regional stability conditions based on the minimization of the cost (2) at each time step that are presented in [5], [6]. Concerning the NMHE approach, local results similar to Theorem 3 are reported in the literature and involve the Newton [19] and Gauss-Newton [18], [29] methods.

The results obtained in the linear case are somehow surprising. The stability conditions for the GMHE and the CGMHE are ensured only for small values of $|A|$, which plays the same role of k_f in the nonlinear case (see Corollaries 1 and 2). By contrast, the application of the Newton method provides a condition for global stability that holds for any $|A|$ via the choice of a sufficiently small μ , as stated in Corollary 3. Note that the NMHE performs as a deadbeat observer if μ is chosen equal to zero.

The stability of the estimation error studied for the GMHE, CGMHE, and NMHE approaches for both linear and nonlinear systems may be easily extended also in case of multiple iterations. For the sake of brevity, let us solely focus on the NMHE for linear systems with M inner steps that are carried out at each time instant instead of only one iteration. The rate of decrease of the estimation error at each time t is much higher. For instance, under the linear assumptions as in (20), we have

$$|e_{t-N}| \leq |x_0 - \bar{x}_0| \left(|A| \left(\frac{\mu}{\mu + \delta_{\min}} \right)^M \right)^{t-N+1}$$

instead of (43), and thus potentially convergent to zero with a faster rate at the price of an increased computational effort.

VII. NUMERICAL RESULTS

The effectiveness of the proposed approaches was investigated with linear and nonlinear examples. The results were compared to those provided by the KF or the EKF as well as to the optimal solution of problem (4) obtained through a “full” optimization, denoted from now as OMHE (optimal MHE). Moreover, in the nonlinear case we considered the iterated EKF [30] for a fair comparison based on fixing the computational time, as will be discussed later. All the simulations were performed in Matlab on a PC equipped with a 2.6 GHz Intel Xeon CPU and 64 GB of RAM. In more detail, the optimization involved in the OMHE was done by using the interior-point algorithm [24] of the routine *fmincon* with default tolerance for the stopping criteria equal to 10^{-9} .

For the purpose of performance evaluation, we used the root mean square error over L simulation runs (RMSE):

$$\text{RMSE}_t := \left(\frac{1}{L} \sum_{l=1}^L |e_t^l|^2 \right)^{1/2}, \quad t = 0, 1, \dots, K-1$$

where e_t^l is the estimation error of the l -th simulation run and K is the simulation length in time steps. Furthermore, we used the asymptotic RMSE (ARMSE):

$$\text{ARMSE} := \frac{1}{S} \sum_{t=K-S}^{K-1} \text{RMSE}_t$$

where S is the backlog length. We considered also the mean of the estimation error over time of the l -th simulation run, i.e.,

$$\bar{e}_l := \frac{1}{K} \sum_{t=0}^{K-1} |e_t^l|, \quad l = 1, \dots, L.$$

To evaluate the computational burden, we accounted for the mean time required to generate a state estimate in the l -th run, denoted by \bar{T}_l .

A. Linear Example

Consider the linear system

$$\begin{aligned} x_{t+1}^{(1)} &= x_t^{(1)} + \Delta t \left(-c_1 x_t^{(1)} + u_t^{(1)} \right) \\ x_{t+1}^{(2)} &= x_t^{(2)} + \Delta t \left(-c_2 x_t^{(2)} + c_1 x_t^{(1)} + u_t^{(2)} \right) \\ x_{t+1}^{(3)} &= x_t^{(3)} + \Delta t \left(-c_3 x_t^{(3)} + c_2 x_t^{(2)} + u_t^{(3)} \right) \\ y_t &= x_t^{(3)} \end{aligned}$$

where $x_t := (x_t^{(1)}, x_t^{(2)}, x_t^{(3)})$, $u_t := (u_t^{(1)}, u_t^{(2)}, u_t^{(3)})$, $c_1 = 1$, $c_2 = 1$, $c_3 = 1$, and sampling time $\Delta t = 0.01$. Such a system may be used to model, for instance, three interconnected reservoirs, with state variables representing the amount of liquid in each basin. Only the level of the last reservoir is measured. In particular, the i -th basin, $i = 1, 2, 3$, has an outflow proportional to $x_t^{(i)}$ (through the coefficient c_i) that enters the $(i+1)$ -th reservoir. Each basin has an external inflow or outflow, represented by the input $u_t^{(i)}$, depending on whether it is positive or negative. Since the state vector represents the amount of liquid in the basins, it is constrained to be non-negative. Moreover, we assume that a maximum capacity exists for each reservoir, and that floodways are present to avoid that the maximum value is exceeded. In more detail, we constrained the state vector x_t to lie in the range $[0, 10]^3$. As regards the inputs u_t , they were randomly generated according to uniform distributions in the interval $[-7, +7]^3$.

We performed $L = 1000$ simulation runs of length $K = 500$ time steps with different initial states randomly chosen according to uniform distributions in the range $[0, 10]^3$. For each trial, the estimate \bar{x}_0 of x_0 was randomly generated by using a Gaussian distribution with mean equal to the real state and three different values for the covariance, i.e., I , $10I$, and $100I$. We choose $S = 50$ to compute the ARMSE.

The parameter μ of the cost (2) was chosen equal to 0.01. The descent step α of the GMHE and CGMHE was taken equal to 0.1. The covariance matrices of the KF were chosen via an experimental tuning. In particular, we fixed them to $10^{-4}I$, and 10^{-3} for the system and measurement noises, respectively, whereas the covariance of the initial estimate was taken equal to the true value. It is worth noting that an analytic solution to the OMHE could be found since the considered system is linear [4]. Instead, we performed the optimization with the interior-point algorithm for the purpose of a fair comparison with the other approaches and with the nonlinear case described in the next section, for which an analytic solution does not exist.

Table I shows the ARMSEs of the various estimation methods together with the medians of $\bar{e} := (\bar{e}_1, \dots, \bar{e}_L)$ and $\bar{T} := (\bar{T}_1, \dots, \bar{T}_L)$ for different horizons N and covariances used to randomly generate \bar{x}_0 . Fig. 3 reports the boxplots of the mean estimation error \bar{e} together with the boxplots of the mean time \bar{T} required to generate a state estimate. Fig. 4 depicts the RMSEs, while Fig. 5 reports the boxplots of the norms of the estimation errors e_t . Lastly, Fig. 6 shows a randomly-extracted trajectory and the corresponding estimates. In Figs. 5 and 6, the results of the CGMHE are omitted since they are similar to the ones of the GMHE. Only the plots with 2 descent iterations are shown for the sake of brevity.

From the simulation results, it turns out that the OMHE guarantees the best estimation accuracy, at the price of the highest computational effort. This is not surprising, as many descent iterations have to be performed in general rather than only few ones as for the GMHE, CGMHE, and NMHE. However, the performances of the NMHE are very similar to the OMHE in terms of ARMSE and \bar{e} , but with much lower computational requirements, which make the NMHE method preferable. The GMHE and CGMHE show about the same estimation accuracy, which is however worse than the one of the NMHE. The gap increases with $\text{cov}(\bar{x}_0)$, i.e., if the difference between x_0 and \bar{x}_0 is large.

As regards the computational time, the KF is the best performer. The GMHE outperforms both the CGMHE and NMHE approaches, as the computations needed to obtain a state estimate are much easier. Moreover, in general there is a slight increase of accuracy if the estimation horizon N grows, at the price of a small increase of the computational effort.

The KF provides the worst values of the ARMSE, as a consequence of the saturation constraints (lower and upper bounds on the state variables). The convergence of the GMHE and CGMHE is quite slow as compared to the NMHE, whereas the RMSEs of the OMHE and NMHE are almost coincident, thus confirming that the NMHE allows for a convenient trade-off between accuracy and computational burden. In general, the larger is the number M of descent iterations the lower is the estimation error, at the price of an increase of the computational effort.

B. Nonlinear Example

Consider a Lorenz chaotic system [31] as follows:

$$\begin{aligned} x_{t+1}^{(1)} &= x_t^{(1)} + \Delta t \left(\sigma x_t^{(2)} - \sigma x_t^{(1)} \right) \\ x_{t+1}^{(2)} &= x_t^{(2)} + \Delta t \left(x_t^{(4)} x_t^{(1)} - x_t^{(1)} x_t^{(3)} - x_t^{(2)} \right) \\ x_{t+1}^{(3)} &= x_t^{(3)} + \Delta t \left(x_t^{(1)} x_t^{(2)} - \beta x_t^{(3)} \right) \\ x_{t+1}^{(4)} &= x_t^{(4)} \\ y_t &= x_t^{(1)} \end{aligned}$$

where $x_t \in \mathbb{R}^4$, $\sigma = 10$, $\beta = 8/3$, and $\Delta t = 0.01$. The state variable $x_t^{(4)}$ is an augmented state variable introduced to estimate the third parameter of the classical Lorenz equation that is supposed unknown [19].

As in the linear example, all the performance indexes were computed over $L = 1000$ simulation runs of length $K = 500$

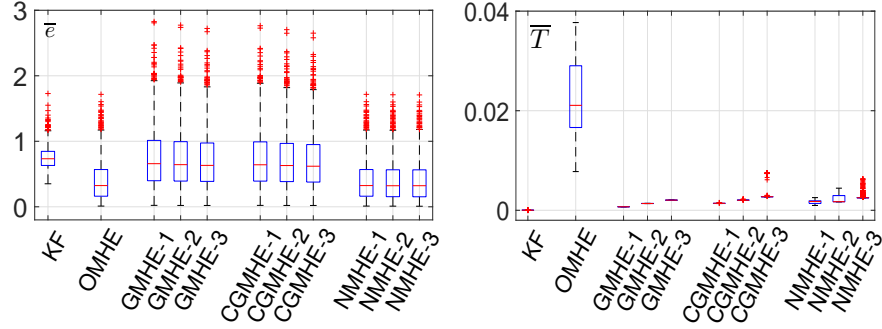


Fig. 3. Mean estimation errors and simulation times (in s) in the linear example with $N = 5$ for the MHE approaches.

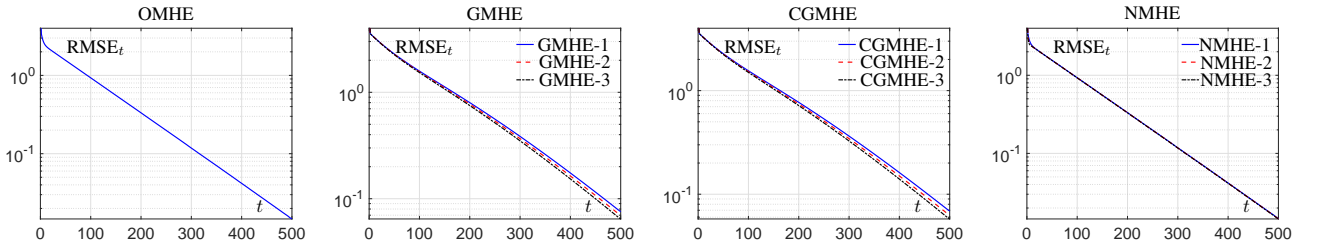


Fig. 4. RMSEs in the linear example with $N = 5$.

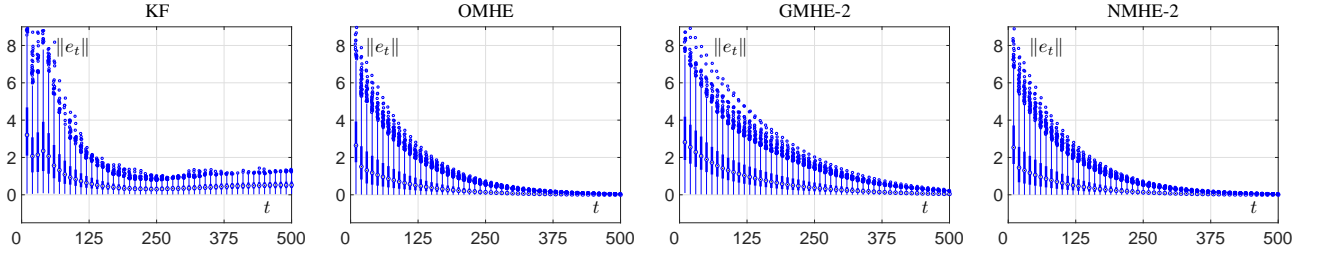


Fig. 5. Norm of the estimation errors in the linear example with $N = 5$ for the MHE approaches.

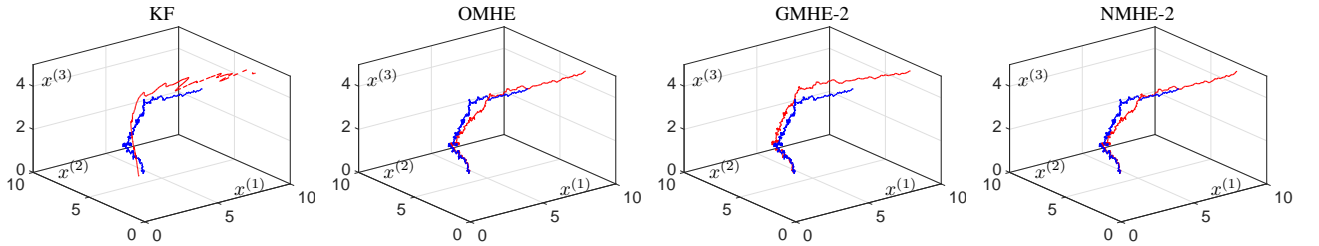


Fig. 6. Trajectory (blue line) and corresponding estimates (red line) in the linear example with $N = 5$ for the MHE approaches.

time steps, and we choose $S = 50$ to compute the ARMSE. For each run, we selected different initial states according to uniform distributions in the ranges $[-10, +10]$ for $x_0^{(1)}$, $x_0^{(2)}$, and $x_0^{(3)}$, and $[0, 30]$ for $x_0^{(4)}$. The estimates \bar{x}_0 of x_0 were randomly generated by using Gaussian distributions with means equal to the real states and three different values for the covariance. In particular, we adopted a covariance equal to I , $10I$, and $100I$ for the first three state variables, whereas the covariance for $x_t^{(4)}$ was taken equal to 1 in all the cases.

As before, the parameter μ of the cost (2) was fixed to 0.01, while the descent step α of the GMHE and CGMHE was taken equal to 0.01. The covariance matrices of the EKF were chosen

again via an experimental tuning. In particular, we fixed them to $10^{-4}I$ and 0.1 for the system and measurement noises, respectively, whereas the covariance for the initial estimate was taken equal to the true one. Similarly to the MHE, in the following we will use the notation EKF- M to denote the iterated EKF with M iteration steps.

Table II reports the ARMSEs of the various estimation methods together with the medians of \bar{e} and \bar{T} . Fig. 7 shows the boxplots of \bar{e} and \bar{T} . Fig. 8 reports the RMSEs, while Fig. 9 contains the boxplots of the norms of the estimation errors e_t . Finally, Fig. 10 shows a randomly-extracted trajectory and the corresponding estimates. In Figs. 9 and 10, the results of the CGMHE are omitted since they are almost the same of the

TABLE I
SIMULATION RESULTS WITH THE LINEAR EXAMPLE FOR DIFFERENT ESTIMATION HORIZONS AND COVARIANCES OF THE INITIAL ESTIMATE.

N	$\text{cov}(\bar{x}_0)$	1	10	100	1	10	100	1	10	100
		ARMSE			median \bar{e}			median \bar{T} [s]		
1	KF	0.59	0.58	0.58	0.52	0.73	1.00	1.93e-5	1.89e-5	1.91e-5
	OMHE	0.02	0.04	0.08	0.21	0.55	1.06	7.32e-3	4.49e-3	4.60e-3
	GMHE-1	0.04	0.11	0.19	0.26	0.68	1.33	1.96e-4	1.91e-4	1.98e-4
	GMHE-2	0.04	0.11	0.19	0.26	0.68	1.32	3.51e-4	3.59e-4	3.55e-4
	GMHE-3	0.04	0.11	0.19	0.26	0.68	1.32	5.04e-4	5.09e-4	5.34e-4
	CGMHE-1	0.04	0.11	0.19	0.26	0.68	1.32	3.59e-4	3.64e-4	3.69e-4
	CGMHE-2	0.04	0.11	0.19	0.26	0.68	1.32	5.18e-4	5.24e-4	5.44e-4
	CGMHE-3	0.04	0.11	0.19	0.26	0.68	1.32	6.87e-4	6.72e-4	7.91e-4
	NMHE-1	0.02	0.04	0.08	0.21	0.55	1.06	5.31e-4	3.06e-4	3.07e-4
	NMHE-2	0.02	0.04	0.05	0.18	0.47	0.94	1.01e-3	5.79e-4	5.83e-4
	NMHE-3	0.02	0.04	0.04	0.16	0.44	0.85	1.48e-3	1.23e-3	1.23e-3
3	OMHE	0.01	0.02	0.04	0.13	0.36	0.73	1.04e-2	1.23e-2	1.00e-2
	GMHE-1	0.04	0.10	0.18	0.26	0.67	1.31	4.25e-4	4.15e-4	4.20e-4
	GMHE-2	0.04	0.10	0.17	0.26	0.67	1.30	8.09e-4	8.14e-4	7.95e-4
	GMHE-3	0.04	0.10	0.17	0.26	0.66	1.29	1.18e-3	1.16e-3	1.20e-3
	CGMHE-1	0.04	0.10	0.17	0.26	0.67	1.30	8.11e-4	7.99e-4	8.19e-4
	CGMHE-2	0.04	0.10	0.17	0.26	0.66	1.30	1.20e-3	1.19e-3	1.22e-3
	CGMHE-3	0.03	0.10	0.17	0.26	0.66	1.29	1.59e-3	1.61e-3	1.57e-3
	NMHE-1	0.01	0.02	0.04	0.13	0.36	0.73	6.03e-4	1.16e-3	1.14e-3
	NMHE-2	0.01	0.02	0.04	0.12	0.34	0.70	1.17e-3	2.18e-3	1.17e-3
	NMHE-3	0.01	0.02	0.04	0.12	0.33	0.69	2.51e-3	3.09e-3	1.73e-3
5	OMHE	0.01	0.02	0.03	0.12	0.32	0.67	2.08e-2	2.11e-2	2.46e-2
	GMHE-1	0.03	0.09	0.16	0.25	0.66	1.28	7.23e-4	7.04e-4	7.12e-4
	GMHE-2	0.03	0.09	0.15	0.25	0.64	1.26	1.38e-3	1.36e-3	1.40e-3
	GMHE-3	0.03	0.08	0.14	0.24	0.63	1.24	2.04e-3	2.04e-3	2.05e-3
	CGMHE-1	0.03	0.09	0.15	0.25	0.64	1.26	1.40e-3	1.40e-3	1.42e-3
	CGMHE-2	0.03	0.08	0.14	0.24	0.63	1.23	2.07e-3	2.02e-3	2.08e-3
	CGMHE-3	0.03	0.07	0.13	0.24	0.62	1.21	2.81e-3	2.73e-3	2.79e-3
	NMHE-1	0.01	0.02	0.03	0.12	0.32	0.67	8.83e-4	1.77e-3	8.87e-4
	NMHE-2	0.01	0.02	0.03	0.12	0.32	0.67	1.71e-3	1.73e-3	1.81e-3
	NMHE-3	0.01	0.02	0.03	0.12	0.32	0.66	3.69e-3	2.52e-3	3.78e-3

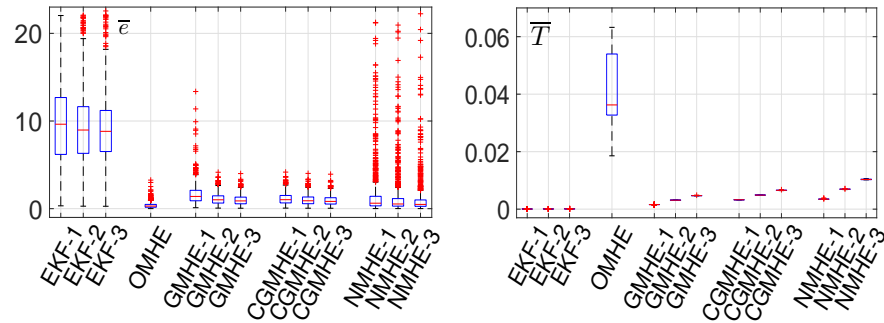


Fig. 7. Mean estimation errors and simulation times (in s) in the nonlinear example with $N = 5$ for the MHE approaches.

GMHE. Only the results obtained with 2 descent iterations are showcased due to space limitations.

From the simulation results, it turns out that the OMHE provides the best results in terms of RMSE and ARMSE, as in the linear example. Unfortunately, it is the most computationally demanding method since it involves the “full” minimization of the cost (2). As compared to the OMHE, the proposed GMHE, CGMHE, and NMHE approaches require much less effort at the price of some decay of performance. Their estimation accuracy increases with the number of descent iterations M , but of course also the computational burden grows with M . If the initial estimate is chosen near the actual state, i.e., if $\text{cov}(\bar{x}_0)$ is small, the NMHE performs better than the GMHE and the CGMHE in terms of estimation error, but its computational effort is higher due to the need of computing

the Hessian matrix of the cost function and its inverse. On the contrary, if the difference between x_0 and \bar{x}_0 grows, the NMHE experiences a decay in the estimation accuracy, and diverging behaviors occur (see the last three columns in Table II). The divergent trajectories are dropped in Figs. 7-9 for clarity of exposition. This is confirmed also by the large presence of outliers in the boxplots of the estimation errors in Fig. 9 for the NMHE, whereas a lower number of outliers may be observed for the other methods. This is also in accordance with the local stability result proved for the NMHE in the nonlinear case as claimed in Theorem 3. By contrast, both the GMHE and CGMHE exhibit a convergent behavior for all the initial states, thus confirming the global stability properties proved in Theorems 1 and 2, respectively. These results highly differ from those of the linear example, where

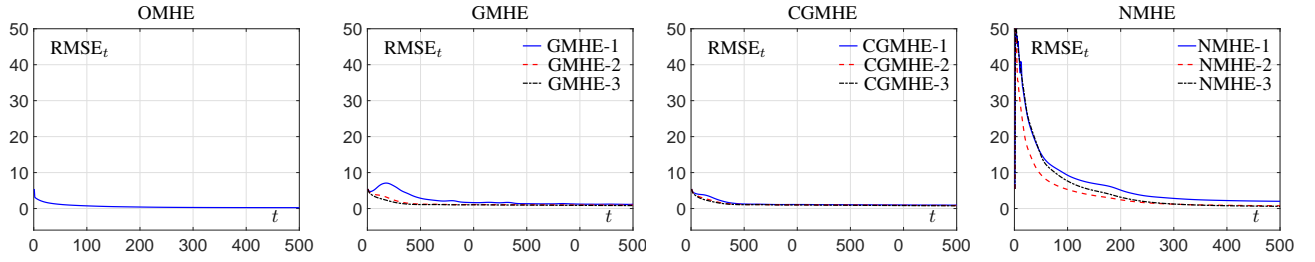


Fig. 8. RMSEs in the nonlinear example with $N = 5$.

TABLE II
SIMULATION RESULTS WITH THE NONLINEAR EXAMPLE FOR DIFFERENT ESTIMATION HORIZONS AND COVARIANCES OF THE INITIAL ESTIMATE.

	cov(\bar{x}_0)	1	10	100	1	10	100	1	10	100	1	10	100
N		ARMSE			median $\bar{\epsilon}$			median \bar{T} [s]			num. of divergences		
1	EKF-1	11.40	11.55	11.67	9.23	9.66	10.21	4.09e-5	4.10e-5	4.09e-5	0	0	0
	EKF-2	10.61	10.62	10.63	8.76	8.97	9.30	6.18e-5	6.16e-5	6.19e-5	0	0	0
	EKF-3	10.52	10.53	10.53	8.60	8.83	9.14	8.32e-5	8.28e-5	8.27e-5	0	0	0
	OMHE	1.01	1.01	1.01	0.68	0.77	1.10	2.82e-3	4.67e-3	2.79e-3	0	0	0
	GMHE-1	8.19	9.80	11.44	2.16	3.66	6.70	5.93e-5	3.53e-5	5.88e-5	0	0	0
	GMHE-2	3.28	5.00	7.35	1.50	2.46	4.67	9.53e-5	5.84e-5	9.63e-5	0	0	0
	GMHE-3	1.93	2.30	4.43	1.26	1.94	3.84	1.32e-4	7.98e-5	1.31e-4	0	0	0
	CGMHE-1	3.35	5.20	7.47	1.52	2.48	4.72	1.08e-4	6.64e-5	1.07e-4	0	0	0
	CGMHE-2	1.98	2.37	4.26	1.26	1.95	3.92	1.52e-4	9.40e-5	1.54e-4	0	0	0
	CGMHE-3	1.50	1.87	2.34	1.15	1.72	3.44	1.96e-4	1.22e-4	2.34e-4	0	0	0
3	NMHE-1	1.01	1.01	1.01	0.68	0.77	1.10	8.86e-5	1.49e-4	1.27e-4	0	0	0
	NMHE-2	0.99	0.99	0.99	0.68	0.76	1.14	1.51e-4	1.55e-4	2.34e-4	0	0	0
	NMHE-3	0.99	0.99	0.99	0.67	0.77	1.17	3.30e-4	2.22e-4	3.34e-4	0	0	0
	OMHE	0.27	0.31	0.52	0.24	0.40	1.00	3.58e-3	3.66e-3	5.99e-3	0	0	0
	GMHE-1	2.20	3.27	4.60	1.32	2.14	4.03	1.04e-4	5.42e-5	5.49e-5	0	0	0
	GMHE-2	1.29	1.30	1.30	1.01	1.44	2.70	1.71e-4	8.84e-5	8.97e-5	0	0	0
	GMHE-3	1.21	1.22	1.20	0.90	1.21	2.25	2.42e-4	1.24e-4	1.23e-4	0	0	0
	CGMHE-1	1.30	1.31	1.32	1.03	1.46	2.75	1.86e-4	1.01e-4	1.01e-4	0	0	0
	CGMHE-2	1.22	1.22	1.20	0.91	1.23	2.28	2.68e-4	2.30e-4	1.44e-4	0	0	0
	CGMHE-3	1.18	1.18	1.16	0.86	1.12	2.05	3.48e-4	2.95e-4	1.83e-4	0	0	0
5	NMHE-1	0.32	1.65	8.23	0.25	0.62	3.06	1.54e-4	1.58e-4	1.56e-4	0	0	12
	NMHE-2	0.27	1.35	6.02	0.21	0.53	2.44	4.22e-4	2.90e-4	2.87e-4	0	0	15
	NMHE-3	0.24	1.16	4.93	0.20	0.49	2.17	6.17e-4	4.23e-4	4.21e-4	0	0	11
	OMHE	0.15	0.24	0.48	0.15	0.30	0.84	4.31e-2	3.63e-2	3.36e-2	0	0	0
	GMHE-1	1.17	1.19	1.16	0.98	1.39	2.58	2.29e-3	1.60e-3	1.59e-3	0	0	0
	GMHE-2	0.96	0.94	0.83	0.77	1.01	1.86	5.42e-3	3.13e-3	3.12e-3	0	0	0
	GMHE-3	0.84	0.84	0.73	0.69	0.89	1.57	4.74e-3	4.71e-3	4.68e-3	0	0	0
	CGMHE-1	0.97	0.95	0.84	0.78	1.02	1.89	5.39e-3	3.26e-3	3.22e-3	0	0	0
	CGMHE-2	0.86	0.85	0.74	0.70	0.90	1.60	4.92e-3	4.89e-3	4.91e-3	0	0	0
	CGMHE-3	0.78	0.78	0.69	0.66	0.83	1.46	9.24e-3	6.55e-3	6.55e-3	0	0	0

the NMHE always outperforms GMHE and CGMHE in terms of estimation accuracy and shows no divergence (see Table I).

For the purpose of comparison, in Table II consider for instance the columns corresponding to the simulation results obtained for initial estimates having covariance equal to 10. It is straightforward to check that EKF-3, CGMHE-2 with $N = 1$, and GMHE-2 with $N = 3$ exhibit quite close medians of \bar{T} , namely $8.28 \cdot 10^{-5}$, $9.40 \cdot 10^{-5}$, and $8.84 \cdot 10^{-5}$, respectively. Indeed, the performances in terms of both ARMSE and median of $\bar{\epsilon}$ are very different: 10.53 and 8.83 for EKF-3; 2.37 and 1.95 for CGMHE-2 with $N = 1$; 1.30 and 1.44 for GMHE-2 with $N = 3$. In other words, the computational effort to apply EKF-3 is about the same of CGMHE-2 with $N = 1$ and GMHE-2 with $N = 3$ but with quite poor performances, especially at steady state, with a difference of about 80%.

VIII. CONCLUSIONS

Three novel, numerically tractable approaches to fast moving horizon state estimation based on the gradient, conjugate gradient, and Newton methods have been described, all provided with a rigorous stability analysis of the estimation error both in the linear and nonlinear settings. The practicability of such methods has been experimentally demonstrated with simulations, also in comparison with the KF/EKF. The simulation results show a behavior concordant with the theoretical findings. The best performances in terms of tradeoff between computational complexity and accuracy are achieved by the Newton method in the linear example and by the gradient or conjugate gradient methods in the nonlinear example. However, the stability conditions required by the gradient and conjugate gradient methods turn out to be quite conservative, thus suggesting to investigate tighter bounds for deriving more

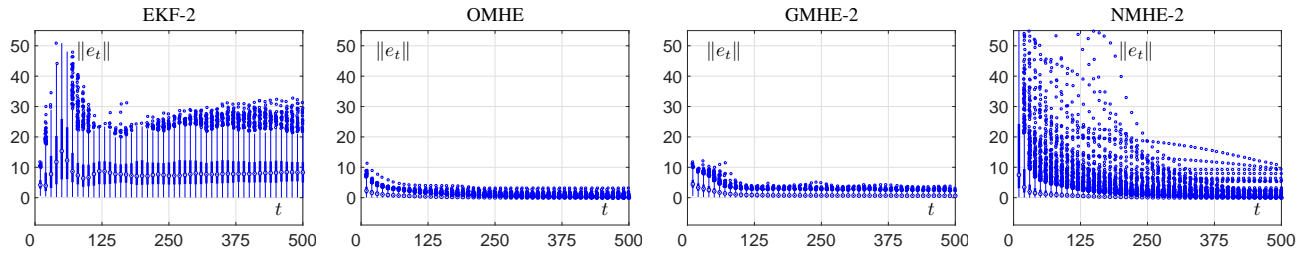


Fig. 9. Norm of the estimation errors in the nonlinear example with $N = 5$ for the MHE approaches.

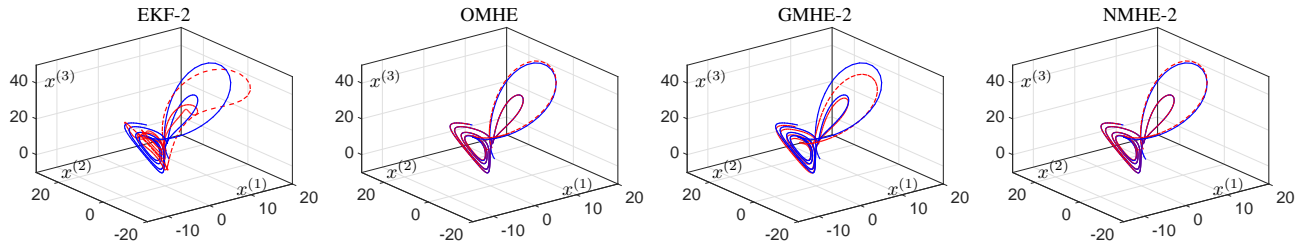


Fig. 10. Trajectory (blue line) and corresponding estimates (red line) in the nonlinear example with $N = 5$ for the MHE approaches.

relaxed conditions. Other descent algorithms will be studied as well, such as the inexact-Newton or other fast methods, to understand the potential of their application to moving horizon state estimation.

REFERENCES

- [1] A. Jazwinski, "Limited memory optimal filtering," *IEEE Trans. Automat. Control*, vol. 13, no. 5, pp. 558–563, 1968.
- [2] C. Rao, J. Rawlings, and J. Lee, "Constrained linear estimation—a moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, 2001.
- [3] C. Rao, J. Rawlings, and D. Mayne, "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations," *IEEE Trans. Automat. Control*, vol. 48, no. 2, pp. 246–257, 2003.
- [4] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for discrete-time linear systems," *IEEE Trans. Automat. Control*, vol. 48, no. 3, pp. 473–478, 2003.
- [5] A. Alessandri, M. Baglietto, and G. Battistelli, "Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes," *Automatica*, vol. 44, no. 7, pp. 1753–1765, 2008.
- [6] A. Alessandri, M. Baglietto, G. Battistelli, and M. Gaggero, "Moving-horizon state estimation for nonlinear systems using neural networks," *IEEE Trans. Neural Networks*, vol. 22, no. 5, pp. 768–780, 2011.
- [7] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "Moving-horizon partition-based state estimation of large-scale systems," *Automatica*, vol. 46, no. 5, pp. 910–918, 2010.
- [8] A. Haber and M. Verhaegen, "Moving horizon estimation for large-scale interconnected systems," *IEEE Trans. Automat. Control*, vol. 58, no. 11, pp. 2834–2847, 2013.
- [9] R. Schneider, R. Hannemann-Tams, and W. Marquardt, "An iterative partition-based moving horizon estimator with coupled inequality constraints," *Automatica*, vol. 61, pp. 302–307, 2015.
- [10] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for switching discrete-time linear systems," *IEEE Trans. Automat. Control*, vol. 50, no. 11, pp. 1736–1748, 2005.
- [11] Y. Guo and B. Huang, "Moving horizon estimation for switching nonlinear systems," *Automatica*, vol. 49, no. 11, pp. 3270–3281, 2013.
- [12] A. Al-Matouq and T. Vincent, "Multiple window moving horizon estimation," *Automatica*, vol. 53, pp. 264–274, 2015.
- [13] A. Alessandri, M. Baglietto, and G. Battistelli, "Min-max moving-horizon estimation for uncertain discrete-time systems," *SIAM J. Control and Optimization*, vol. 50, no. 3, pp. 1439–1465, 2012.
- [14] L. Fagiano and C. Novara, "A combined moving horizon and direct virtual sensor approach for constrained nonlinear estimation," *Automatica*, vol. 49, no. 1, pp. 193–199, 2013.
- [15] A. Alessandri and M. Awawdeh, "Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers," *Automatica*, vol. 67, pp. 85–93, 2016.
- [16] Y. Wan, T. Keviczky, M. Verhaegen, and F. Gustafsson, "Data-driven robust receding horizon fault estimation," *Automatica*, vol. 71, pp. 210–221, 2016.
- [17] A. Alessandri and M. Gaggero, "Moving-horizon estimation for discrete-time linear and nonlinear systems using the gradient and Newton methods," in *55th IEEE Conf. on Decision and Control*, Las Vegas, Nevada, USA, 2016, pp. 2906–2911.
- [18] E. Bıyık and M. Arcak, "A hybrid redesign of Newton observers in the absence of an exact discrete-time model," *Systems & Control Letters*, vol. 55, no. 6, pp. 429–436, 2006.
- [19] A. Wynn, M. Vukov, and M. Diehl, "Convergence guarantees for moving horizon estimation based on the real-time iteration scheme," *IEEE Trans. Automat. Control*, vol. 59, no. 8, pp. 2215–2221, 2014.
- [20] V. Zavala, C. Laird, and L. Biegler, "A fast moving horizon estimation algorithm based on nonlinear programming sensitivity," *Journal of Process Control*, vol. 18, no. 9, pp. 876–884, 2008.
- [21] P. Kühl, M. Diehl, T. Kraus, J. Schlöder, and H. Bock, "A real-time algorithm for moving horizon state and parameter estimation," *Computers & Chemical Engineering*, vol. 35, no. 1, pp. 71–83, 2011.
- [22] H. Jang, J. Lee, R. Braat, and K.-K. Kim, "Fast moving horizon estimation for a two-dimensional distributed parameter system," *Computers & Chemical Engineering*, vol. 63, no. 1, pp. 159–172, 2014.
- [23] B. Morabito, M. Kögel, E. Bullinger, G. Pannocchia, and R. Findeisen, "Simple and efficient moving horizon estimation based on the fast gradient method," in *Proc. 5th IFAC Conference on Nonlinear Model Predictive Control*, Seville, Spain, 2015, pp. 428–433.
- [24] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
- [25] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *Computer Journal*, vol. 7, pp. 149–154, 1964.
- [26] R. Dembo, S. Eisenstat, and T. Steihaug, "Inexact Newton methods," *SIAM Journal on Numer. Anal.*, vol. 19, no. 2, pp. 400–408, 1982.
- [27] C. T. Kelley, *Iterative Methods for Optimization*, ser. Frontiers in Applied Mathematics. Philadelphia, PA: SIAM, 1999, vol. 18.
- [28] S. Hanba, "Further results on the uniform observability of discrete-time nonlinear systems," *IEEE Trans. Automat. Control*, vol. 55, no. 4, pp. 1034–1038, 2010.
- [29] P. Moraal and J. Grizzle, "Observer design for nonlinear systems with discrete-time measurements," *IEEE Trans. Automat. Control*, vol. 40, no. 3, pp. 395–404, 1995.
- [30] A. Gelb, *Applied Optimal Estimation*. M.I.T. Press, 1974.
- [31] E. Lorenz, "Deterministic nonperiodic flow," *Journal of Atmospheric Sciences*, vol. 20, pp. 130–141, 1963.



Angelo Alessandri received the “Laurea” degree in electronics engineering and the Ph.D. degree in electronics engineering and computer science from the University of Genoa, Italy, in 1992 and 1996, respectively. He was a Visiting Scientist with the Naval Postgraduate School, Monterey, CA, in 1998. From 1996 to 2005, he was a Research Scientist with the National Research Council of Italy, Italy. In 2005, he joined the University of Genoa, where he is currently an Associate Professor. His current research interests include optimal control, linear and

nonlinear estimation, and learning applications in automatic control. Dr. Alessandri was an Associate Editor of the IEEE Transactions on Control Systems Technology, the IEEE Transactions on Neural Networks, and the IFAC Journal of Engineering Applications of Artificial Intelligence. He is an Editor of the International Journal of Adaptive Control and Signal Processing. He is an Associate Editor of the IEEE Control Systems Society Conference Editorial Board.



Mauro Gaggero received the B.S. and M.S. degrees in electronics engineering and the Ph.D. degree in mathematical engineering from the University of Genoa, Italy, in 2003, 2005, and 2010, respectively. He was a Post-Doctoral Fellow with the Faculty of Engineering, University of Genoa, in 2010. Since 2011, he has been a Research Scientist with the Institute of Intelligent Systems for Automation, National Research Council of Italy, Italy. His current research interests include control, estimation, and optimization of nonlinear systems, distributed parameter systems, neural networks, and functional optimization problems. Dr. Gaggero is an Associate Editor of the European Control Association Conference Editorial Board and of the IEEE Control Systems Society Conference Editorial Board.

nonlinear estimation, and learning applications in automatic control. Dr. Alessandri was an Associate Editor of the IEEE Transactions on Control Systems Technology, the IEEE Transactions on Neural Networks, and the IFAC Journal of Engineering Applications of Artificial Intelligence. He is an Editor of the International Journal of Adaptive Control and Signal Processing. He is an Associate Editor of the IEEE Control Systems Society Conference Editorial Board.